DOCUMENT RESUME

ED 054 793		LI 003 068
AUTHOR	Bookstein, Abraham	
TITLE	Congestion at Card and Book Catalogs--A Queuing Theory Approach.	
PUB DATE	71	
NOTE	32p.; (6 References)	
EDRS PRICE	MF-$0.65 HC-$3.29	
DESCRIPTORS	*Book Catalogs; *Catalogs; *Design; *Dictionary Catalogs; *Divided Catalogs; Library Services; Mathematical Models; Models	

ABSTRACT

The question of whether a library's catalog should consist of cards arranged in a single alphabetical order (the "dictionary catalog) or be segregated as a separate file is discussed. Development is extended to encompass related problems involved in the creation of a book catalog. A model to study the effects of congestion at the catalog is created. Using a drawer chosen randomly from either a dictionary catalog, or the subject or author-title part of a split catalog, three measures of congestion are considered: (1) the probability that the drawer is being used, (2) the average time needed to wait for a use and (3) the average number of people attracted to the drawer at any time. All the parameters used and the basic relations among them are collected in Section II. The first measure of congestion considered is the likelihood that a user must wait before he can use a drawer. The next measure of congestion is the mean time a user must wait to gain access to a drawer. The final measure of congestion is the number of people contributed to the system at any time along each drawer. Section VI considers the implications of the model for the construction of book catalogs. It was found that each of the three criteria of congestion can lead to a different conclusion. (Author/NH)

CONGESTION AT CARD AND BOOK CATALOGS — A QUEUING THEORY APPROACH

Abraham Bookstein

CONGESTION AT CARD AND BOOK CATALOGS —

A QUEUING THEORY APPROACH

I. Introduction

The catalog is the most heavily used of a library's files. Valuable to the patron as his main form of access to the library's collection, the catalog also plays a vital role in much of the library's internal processes. As such, the choice of catalog design is deserving of the most careful attention of library planners and managers. We here discuss one question related to catalog design: should a catalog consist of cards arranged in a single alphabetical order (the "dictionary" catalog), or should the subject cards be segregated as a separate file? Our development will then be extended to encompass related problems involved in the creation of a book catalog.

A library manager making such decisions has to take many factors into consideration. He must understand patterns of behavior in the use of the catalog, an area in which much research remains to be done. McGregor,[1] in defense of the dictionary catalog, notes that splitting the catalog would force a number of readers to consult two files to find, for example, all books by and about an author or organization or else lose pertinent material; it would be useful to have statistics regarding what fraction of users would be affected in this manner. The proponents of splitting the catalog point to the confusion resulting from merging the two sets of cards, and the extra effort required of a user who must go through many extraneous cards before reaching the one he is seeking; but it must also be considered that relatively few people understand the logic of the card catalog and thus

there would be some confusion as to which catalog is appropriate for, e.g. an autobiography, or a book with a corporate author.[1]

Another factor that must be considered is cost: the relative costs of establishing the catalog in one form or another, and the relative costs of maintaining it. The library manager must balance these various considerations and make a decision based on his understanding of what will best serve his institutions's objectives. To do this the manager must to the extent possible understand the nature of his problem and the consequences of his decision. Careful research is needed and can be assisted by the creation of analytical models that abstract from the system those characteristics most important for the decision.

It is our purpose to create a model to study the effects of congestion at the catalog, a problem recognized by McGregor and others as important in deciding whether the catalog should be split. The literature presents much conflicting experience on this subject, but there is little attempt at theoretical analysis. Heinritz,[2] following Lubetsky,[3] notes that use of subject headings tends to take a longer time than author-title card use since the user is uncertain as to what item he is looking for, and usually must examine a number of cards, whereas the user of a main entry card need only locate a single card and make a note of the call number. It is thus possible to argue that keeping a dictionary catalog penalizes the user of author-title cards by forcing him to wait until the person making a subject search is finished with the card drawer; on the other hand, as Lubetsky observed, dividing the catalog might "relieve congestion at the author-title catalog, or catalogs, but at the same time seriously aggravate congestion at the subject catalog, where all the subjects will now be compacted in a smaller number of drawers than they occupied in a single

catalog." Heinritz questions this assertion and suggests that parameters such as the relative number of subject drawers as compared to author-title drawers, the mean time required for a use of a subject card and the relative rate of users intending to look for a subject heading, as against an author-title heading will prove to be of value in deciding this question. He offers as a measure of congestion the quantity: the rate at which users of a type arrive, multiplied by the time required for a use and divided by the numbers of drawers.

We intend to extend Heinritz' observation by introducing the effect of variability, a factor found to be of crucial importance in other systems suffering from problems of congestion. That this should be the case can be seen by comparing two simple hypothetical cases. Suppose it is found that each user takes exactly 20 seconds to use a catalog drawer, and that users arrive at exactly 20 second intervals. In such a situation there would be no congestion at each drawer; a patron would arrive just in time to have the desired drawer handed to him for his own use. But consider what may happen if we allow some variation in arrival rate and time required for use, while leaving the average magnitude of these quantities unchanged. Instead of the first user arriving "on time," the drawer may be left unused for 30 seconds. Then the first user arrives, and is quickly followed by users two and three. As luck would have it, the first user keeps the drawer not for 20 seconds, but for 40, and by the time he has finished using it, five people are waiting in line. This pattern, unfortunately, is found to be the rule for systems in which variability is allowed. Accordingly, in our model for catalog congestion, we will take variability into account.

Our approach is that of queuing theory.[4] We consider catalog users as people arriving for a "service" and forming a line in front of the

"server" (the catalog drawer) if it is "busy" (being used). Such a model fits many other processes in a library, ranging from patrons lining up to check out a book, to erroneously typed bibliographic entries for a computerized catalog lining up for correction. However, with the striking exception of Morse's elegant treatment of book circulation,[5] this powerful branch of applied mathematics has not found its way into library literature.

Our intention, then, is to predict the extent of congestion at the catalog on the basis of this model, and to make explicit the parameters that must be measured before this quantity can be estimated. The problem of congestion is associated with such inconveniences as the traffic encountered in making one's way to a desired drawer, having access to a drawer interfered with by other users in the vicinity of the drawer, and finding the desired drawer already in use. We will offer three quantifiable measures of congestion that we believe will be correlated with the above difficulties. We will study a drawer chosen randomly from either a dictionary catalog, or the subject or author-title part of a split catalog and consider the three measures:

1) the probability that the drawer is being used,

2) the average time needed to wait for a use, and

3) the average number of people attracted to the drawer at any time.

The last of these can be related to the number of people in the area of the catalog if one knows the number of drawers in the catalog; the first two measures relate to user convenience. (This model considers only people using the catalog, or waiting for its use; it doesn't consider passers-by or loiterers.) We will compare these measures for the subject and author-title catalogs of a split catalog, and by taking an average of these make an overall comparison of the split and dictionary catalogs.[1] It will be seen that the problem is quite complex, and that the three measures do not always lead to the same conclusions; a library

manager would have to decide which is the most appropriate for his circumstances.

We emphasize that the library manager must be concerned with other considerations in addition to that of congestion, and on the basis of these he may well decide upon a policy that will not minimize congestion. This is being written in the hope that if he should do so, it will be with a realistic understanding of the cost in congestion that he would have to pay for the other advantages. Models such as these cannot in themselves make policy; they can only provide a means by which policy can be made more rationally. The point at which congestion would become so intolerable as to make its alleviation worth some sacrifice of performance of a different nature must be decided by the policy maker.[2]

II. Description of parameters

It will be useful to collect in a single section all the parameters of use in the ensuing sections, as well as the basic relations among them. In the following, those parameters whose values change as we consider different catalogs will be subscripted, with "s" standing for subject catalogs, "a" for author-title catalogs, "d" for dictionary catalogs and "sp" for split catalogs; in such cases, the unsubscripted symbol refers to the generic quantity, a correct relation resulting if all such symbols are subscripted consistently. In the following description we define such symbols by placing an * where a subscript might be placed. We suggest that the reader skim this section to get an idea of what variables are significant for estimating congestion, and then refer to it as needed.

The basic parameters are:

λ_* Rate of arrivals to a drawer. Each member of a set of multiple uses is considered as a separate use;

α Probability that a patron will make a subject heading use. This quantity is estimated by the fraction of users making such a search. $1-\alpha$ will accordingly be the probability of an author-title search;

f The fraction of cards in the drawer being considered that are subject cards. $1-f$ is thus the fraction of cards that are author-title cards. We estimate total congestion at the catalog by using for f the fraction of cards that are subject cards in the total catalog.

t_* The expected time required for a use. Thus $t_s = \int_0^\infty t g_s(t)dt$, for $g_s(t)$ the distribution of time required for a subject use;

v_* The expectation for the square of the time required for a search, i.e. $v_s = \int_0^\infty t^2 g_s(t)dt$. This can be estimated by the average of the squares of time taken for a use by a number of users;

W_* Expected waiting time for use of a drawer;

n_* Expected number of people around a drawer, either using it or waiting for use.

Other symbols that will be convenient are:

N Number of cards in a drawer; if this is a dictionary catalog drawer, fN of these will be subject cards and $(1-f)N$ will

be author-title cards; N is assumed equal for all catalogs;

$\bar{\lambda}_*$ Arrival rate per card;

κ_* Defined as $\frac{v}{2t^2}$. Thus $\kappa_s = \frac{v_s}{2t_s^2} = \frac{1}{2}\left(1 + \frac{\text{variance}_s}{t_s^2}\right)$,

and is a measure of the variability of the distribution;

ρ_* Defined as $\lambda_* t_*$ and related to intensity of use. ρ must always be less than one if the system is to be manageable; this will almost certainly be the case in practice and will be assumed in the following.

It will be useful to derive a number of relations between these parameters. These will relate what can be measured in different catalogs to each other and to what appears in our final results.

Since $\bar{\lambda}_s(fN)$ is the number of subject users arriving at a dictionary catalog drawer in a unit of time, we immediately have

$$1) \qquad \alpha = \frac{\bar{\lambda}_s fN}{\lambda_d} = \frac{\lambda_s f}{\lambda_d} ,$$

since $\lambda_s = \bar{\lambda}_s N$ is the rate of arrivals for a drawer consisting only of subject cards.

We similarly have

$$2) \qquad 1-\alpha = \frac{\lambda_a(1-f)}{\lambda_d}$$

Our assumption that the value of the parameters doesn't depend strongly on the type of catalog implies

$$3) \qquad \lambda_d = \overline{\lambda}_s fN + \lambda_a(1-f)N = \lambda_s f + (1-f)\lambda_a .$$

We observe that $\alpha/f = \lambda_s/\lambda_d$ and $\frac{1-\alpha}{1-f} = \frac{\lambda_a}{\lambda_d}$ are not necessarily equal to one, i.e. the probability of a patron using a subject card is not necessarily equal to the probability of randomly picking a subject card from the catalog. If $\alpha<f$ we describe the subject headings as "underused"; if $\alpha>f$ we consider the subject headings to be "overused." Our t_s and t_a correspond to Heinritz' times for use of a catalog; our λ_a and λ_s correspond to his rate of drawer occupation, and our f and 1-f relate to his parameter giving the number of drawers of one type as compared to the other (since if separated, the fraction of drawers full of subject cards will equal the fraction of cards in the dictionary catalog that are subject cards).

To compare the split catalog and the dictionary catalog, it will be necessary to relate the service times of one to those of the other. We will assume that the overall service distribution for the dictionary catalog is a mixture of the distributions of times for pure subject and pure author-title uses, with α and $1-\alpha$ being the respective weights. We then have $t_d = \alpha t_s + (1-\alpha)t_a$. This result depends on the time needed to use a catalog, if we know it is, for example, a subject use, being independent of whether these cards are alone or mixed with author-title cards. This may or may not be the case, since it is quite possible that merging the two types of cards will slow card search because of an increase of confusion resulting from having a greater number of cards in an interval of the alphabet; this aspect of catalog use bears investigation. We feel it is reasonable that

such an effect, if present, is small compared to the total process of catalog use. We similarly assume that changing from one form of catalog to another will not seriously affect the arrival rate by, for example, discouraging users in a significant manner. These assumptions on the integrity of the parameters will be used throughout the paper, though the model can be modified to include these effects.

From this point on we will make the assumption, verified in a great number of similar situations, that the user arrivals can be adequately approximated by a Poisson distribution. By this we mean that arrivals are assumed to be independent of each other and that at any instant the probability of an arrival occurring does not depend on the time that elapsed since the previous arrival. If it is found that arrivals significantly cluster about certain parts of the day, perhaps between classes, or at lunch hour, it may be desired to make the calculations using parameters measured both at peak and lax periods.

III. Probability of Blocking

As our first measure of congestion we consider the likelihood that a user must wait before he can use a drawer, that is, that his use is blocked. We use the result of single server systems that, for Poisson arrivals, this probability is given by the product of the arrival rate and the expected time of use. For a split catalog, we denote by $\rho_a = \lambda_a t_a$ and $\rho_s = \lambda_s t_s$ probabilities that an author-title user or a subject user will find this drawer in use. If ρ_{sp} denotes the probability of a random user of a split catalog finding his drawer in use, we have:

$$1) \qquad \rho_{sp} = \alpha\rho_s + (1-\alpha)\rho_a .$$

On the other hand, since $\alpha t_s + (1-\alpha)t_a$ is the expected service time at a dictionary catalog, $\rho_d = \lambda_d[\alpha t_s + (1-\alpha)t_a]$. Substituting the values of α and $1-\alpha$ derived in part II, we conclude:

$$2) \qquad \rho_d = f(\lambda_s t_s) + (1-f)(\lambda_a t_a) = f\rho_s + (1-f)\rho_a.$$

We can now make a number of observations:

a) If $\rho_s = \rho_a = \rho$, then also $\rho_{sp} = \rho_d = \rho$, since, e.g.

$$\rho_{sp} = \alpha\rho_s + (1-\alpha)\rho_a = \alpha\rho + (1-\alpha)\rho = [\alpha + (1-\alpha)]\rho = \rho.$$

In such a case, removing subject cards from the dictionary catalog will not have any effect on congestion. This condition can be stated as $\lambda_s t_s = \lambda_a t_a$, or $(\frac{\alpha}{f}\lambda_d)t_s = (\frac{1-\alpha}{1-f}\lambda_d)t_a$, which becomes $\frac{\alpha}{f}t_s = \frac{1-\alpha}{1-f}t_a$. The parameters in the last form can easily be measured at a dictionary catalog; those of the first at a split catalog. If $\alpha=f$, this condition becomes $t_s = t_a$, the equality of service time for each type of user. If $t_s > t_a$, however, we may still have equality if the subject cards are underused by the appropriate amount.

b) If $\alpha = f$, then $\rho_{sp} = \rho_d$, though here ρ_s need not be the same at ρ_a. If, say, $\rho_s > \rho_a$, we have $\rho_s > \rho_d = \rho_{sp} > \rho_a$.

c) Suppose $\rho_s > \rho_a$ (i.e., $\frac{\alpha}{f}t_s > \frac{1-\alpha}{1-f}t_a$). Then $\rho_d = \rho_a + f(\rho_s - \rho_a) > \rho_a$ since $f(\rho_s - \rho_a) > 0$. Similarly $\rho_d = (1-(1-f))\rho_s + (1-f)\rho_a = \rho_s + (1-f)(\rho_a - \rho_s) < \rho_s$ (recall $f<1$). Thus $\rho_s > \rho_d > \rho_a$.

This states that if the subject cards are in fact causing congestion, that is, if they would produce more congestion by themselves than the author-title cards would, then they will also have more congestion than the dictionary catalog. This is equivalent to saying that if we have a dictionary catalog and its congestion could be reduced by removing the subject cards, then we will <u>always</u> have a compensating increase of congestion at the subject catalog; the author-title users will be relieved only at the expense of the subject users.

d) If $\rho_a > \rho_s$, the analysis given in part b) is valid here, but the inequalities are reversed.

We can finally ask when the dictionary catalog will be preferred to the split catalog. This will be the case when $f\rho_s + (1-f)\rho_a < \alpha\rho_s + (1-\alpha)\rho_a$. We conclude that a dictionary catalog will be preferred when

3) $$(f - \alpha)(\rho_s - \rho_a) < 0.$$

If the inequality 3) is not satisfied, a split catalog is preferred.

If $\alpha < f$ (subject cards underused), then a dictionary catalog will be preferred when $\rho_s < \rho_a$, i.e. $\frac{\alpha}{f} t_s < \frac{1-\alpha}{1-f} t_a$. If not, it will be preferable to split the catalog.

If $\alpha < f$ (subject cards overused), then a dictionary catalog will be preferred if $\rho_s > \rho_a$, i.e. $\frac{\alpha}{f} t_s > \frac{1-\alpha}{1-f} t_a$; otherwise the split catalog is preferred.

For public libraries, where it is likely that $\alpha > f$ and almost certainly $\frac{\alpha}{f} t_s > \frac{1-\alpha}{1-f} t_a$, from the standpoint of this measure of congestion alone a dictionary catalog will be preferred. For research

libraries, where $\alpha < f$ seems usual and $\frac{\alpha}{f} t_s > \frac{1-\alpha}{1-f} t_a$ is still likely, it would be preferable to split the catalog. As we noted at the beginning of this section, the results are very general, since it is very likely that arrivals are Poisson in nature.

IV. Waiting Time

The next measure of congestion to be considered will be the mean time a catalog user must wait before he can gain access to a drawer. We will assume that service times are independent of each other; the Poisson property assumed of arrivals, though likely valid, will not be necessary for service times. With this assumption the Pollaczek-Khintchine formula [4]

1) $$W = \frac{\lambda v}{2[1-\lambda t]} ,$$

is valid. Since $v = \int_0^\infty t^2 f(t)dt$ and $f_d(t) = \alpha f_s(t) + (1-\alpha)f_a(t)$, we conclude

2) $$v_d = \alpha v_s + (1-\alpha)v_a .$$

Thus we find

3) $$W_s = \frac{\lambda_s v_s}{2[1-\lambda_s t_s]} = \frac{\lambda_s \kappa_s t_s^2}{[1-\rho_s]} = \frac{\frac{1}{\lambda_s} \kappa_s \rho_s^2}{[1-\rho_s]} ,$$

where we use $\frac{1}{2} v_s = \frac{t_s^2}{2} \frac{v_s}{t_s^2} = t_s^2 \kappa_s$. Similarly

4) $$W_a = \frac{\lambda_a v_a}{2[1-\lambda_a t_a]} = \frac{\lambda_a \kappa_a t_a^2}{[1-\rho_a]} = \frac{\frac{1}{\lambda_a} \kappa_a \rho_a^2}{[1-\rho_a]} .$$

Using our expressions for v_d and t_d we get

$$W_d = \frac{1}{2}\;\frac{\lambda_d[\alpha v_s + (1-\alpha)v_a]}{1-\lambda_d[\alpha t_s + (1-\alpha)t_a]} = \frac{1}{2}\;\frac{f\lambda_s v_s + (1-f)\lambda_a v_a}{1-[f\lambda_s t_s + (1-f)\lambda_a t_a)} =$$

$$\frac{\frac{f}{\lambda_s}(\lambda_s t_s)^2 \frac{v_s}{t_s^2} + \frac{(1-f)}{\lambda_a}(\lambda_a t_a)^2 \frac{v_a^2}{t_a^2}}{1-[\lambda_s f t_s + (1-f)\lambda_a t_a]}$$

Thus

$$5) \qquad W_d = \frac{\frac{f}{\lambda_s}\kappa_s\,\rho_s^2 + \frac{(1-f)}{\lambda_a}\kappa_a\,\rho_a^2}{1-(f\rho_s + (1-f)\rho_a)}$$

Since $W_{sp} = \alpha W_s + (1-\alpha)W_a$, we have

$$6) \quad W_{sp} = \frac{\frac{\alpha}{\lambda_s}\kappa_s\,\rho_s^2}{1-\rho_s} + \frac{\frac{1-\alpha}{\lambda_a}\kappa_s\,\rho_a^2}{1-\rho_a} = \frac{\frac{f}{\lambda_d}\kappa_s\,\rho_s^2}{1-\rho_s} + \frac{\frac{1-f}{\lambda_d}\kappa_s\,\rho_a^2}{1-\rho_a}\,.$$

It is noted that κ plays an important role in these equations, and it is through κ that variability influences the system. If the time required for subject and author-title use were Poisson processes, then κ_s and κ_a would be one; we suspect this will in fact be the case, though we will continue to include the κ's in the equations to maintain generality, and to make explicit the influence of variability. Though we nowhere in our analyses will need $\kappa_a = \kappa_s = 1$, it will at points be useful if $\kappa_s = \kappa_a$, and at those points we will make that assumption. We note here that only

two moments of the distribution are needed to determine the effect on congestion!

The equations are defined in terms of quantities that can be measured at either a dictionary or a split catalog, and provide a means by which a manager may estimate the effect of changing the form of his catalog.

We observe that once again it is impossible to relieve congestion at a dictionary catalog by, for example, removing subject cards, without increasing congestion at the subject catalog thus produced. (A similar conclusion follows, if it is the author-title cards that are causing congestion.) The assertion is proved in Appendix I.

We can now examine special, interesting cases:

a) $\rho_a = \rho_s$, $\alpha = f$, $\kappa_a = \kappa_s$.

Here we immediately find $W_a = W_s = W_d = W_{sp}$.

b) $\rho_a = \rho_s$, $\alpha = f$. Since $\alpha = \frac{\lambda_s}{\lambda_d} f$ and $(1-\alpha) = \frac{\lambda_a}{\lambda_d}(1-f)$, $\alpha = f$ implies $\lambda_s = \lambda_a = \lambda_d$; thus $W_{sp} = W_d$

and there is no basis to prefer one form of catalog to the other. It is no longer the case that $W_s = W_a$, however; that part of the split catalog with the smaller κ will produce smaller waiting times than the dictionary catalog at the expense of the other part. It is interesting that while merely equating ρ_a to ρ_s produced equal congestion in all catalogs, if we use the probability of blocking as the measure of congestion, this is

now no longer the case; it is possible that the probability of blocking is the same for the two components of a split catalog while the waiting times differ -- indeed, because of the continuity of the expressions, it is possible that the subject catalog may perform better than the author-title catalog by one measure, and worse by the other. We do note, however, that κ_s is likely to be near κ_a.

c) $\rho_a = \rho_s = \rho$

The relation between W_s and W_a is the same as that between $\frac{\kappa_s}{\lambda_s}$ and $\frac{\kappa_a}{\lambda_a}$;

i.e. $W_s > W_a$ if $\frac{\kappa_s}{\lambda_s} > \frac{\kappa_a}{\lambda_a}$ or $\frac{\kappa_s}{\kappa_a} > \frac{\lambda_s}{\lambda_a} = \frac{\alpha}{1-\alpha} \frac{1-f}{f}$.

We now ask when a split catalog has better overall performance than a dictionary catalog, that is, when is $W_{sp} < W_d$. It is shown in Appendix II that for $\rho_s = \rho_a$, this will always be so if $\kappa_s = \kappa_a$.

d) $\alpha = f$

If $\kappa_s = \kappa_a$, this condition implies that a dictionary catalog is always to be preferred (see Appendix III). Thus we see, if we start with $\rho_s = \rho_a$ and $\alpha = f$, increasing the difference between ρ_a and ρ_s produces the opposite effect of increasing the difference between α and f. If these differences are large, one must make an exact calculation to see the effect; if they are small and one is significantly larger than the other, the smaller quantity can effectively be set equal to zero. Again, on the basis of continuity of the expressions, it is conceivable that a dictionary catalog will be preferred on the basis of the blocking probability

criterion, and a split catalog preferred on the basis of waiting times. To see this, suppose $\alpha = f$ and $\rho_s > \rho_a$. By the waiting time criterion, a dictionary catalog is preferred. Now increase f very slightly, so that W_d is still less than W_{sp}. For this case, however, the blocking criterion will prefer a split catalog.

For the final special case we examine this effect by choosing values $\rho_a \neq \rho_s$, $\alpha \neq f$ for which we can perform the analysis,

e) $\rho_a << \rho_s$.

We assume here that t_a is so small that we can ignore ρ_a and $\kappa_a \rho_a$ as compared to the corresponding subject parameters. This effectively yields

$$W_{sp} = \frac{\kappa_s f \rho_s^2}{1-\rho_s} \quad \text{and} \quad W_d = \frac{\kappa_s f \rho_s^2 \left(\frac{f}{\alpha}\right)}{1-f\,\rho_s} .$$

Thus $W_{sp} > W_d$ if $\frac{1}{1-\rho_s} > \frac{\frac{f}{\alpha}}{1-f\rho_s}$, which is equivalent to $\rho_s > \frac{1}{f}\frac{f-\alpha}{1-\alpha}$.

If $f = \alpha$, we prefer a dictionary catalog, a result that agrees with our previous analysis. We also note that $f(1-\alpha) < 1$, so in this region, $\rho_s = (\rho_s - \rho_a)$ must be increased <u>more</u> than $(f-\alpha)$ to keep it advantageous to a dictionary catalog. Here $(\rho_s - \rho_a) > (f-\alpha)$ would not suffice.

V. Number of People in the System

The final measure of congestion is the number of people contributed to the system at any time by each drawer, exclusive of loiterers and passersby. Our assumptions are identical to those in part IV. We make use of a very general conservation law, $E(n) = \lambda E(t)$, that relates $E(n)$, the expected number of people in the system, to $E(t)$, the expected

processing time. If the line is considered as the system, $n_L = \lambda W$, for n_L the number of people in line, and W the time spent on line. Since λt is the expected number of people using the drawer, the number of people around the drawer is $n = n_L + \lambda t$.

Applying this equation to the different catalogs yields:

1) $$n_s = \frac{\kappa_s \rho_s^2}{1-\rho_s} + \rho_s,$$

2) $$n_a = \frac{\kappa_a \rho_a^2}{1-\rho_a} + \rho_a,$$

3) $$n_d = \frac{\kappa_s f(\frac{f}{\alpha})\ \rho_s^2 + (1-f)\ \kappa_a \frac{(1-f)}{1-\alpha}\ \rho_a^2}{1-(f\rho_s + (1-f)\ \rho_a)} + f\rho_s + (1-f)\rho_a;$$

4) $$n_{sp} = \frac{\kappa_s f \rho_s^2}{1-\rho_s} + \frac{\kappa_a (1-f)\ \rho_a^2}{1-\rho_a} + f\rho_s + (1-f)\rho_a.$$

(In equation 4) we use f and $1-f$ as weights instead of α and $1-\alpha$ because we are now selecting a random drawer rather than a random arrival.)

The analysis comparing n_{sp} and n_d is identical to that comparing W_{sp} and W_d, since the expected number of people using a drawer does not distinguish between n_{sp} and n_d, and the remaining term is, within a factor of $1/\lambda_d$, the same as the corresponding terms for W. Our conclusions are accordingly the same as those in part IV.

A major difference appears, however, when we compare n_s, n_a and n_d. This difference can be seen clearly if we examine the special case where $\kappa_s = \kappa_a = \kappa$ and $\rho_s = \rho_a = \rho$. Here $n_a = n_s = \frac{\kappa\rho^2}{1-\rho} + \rho$, and

$$n_d = \frac{\kappa\rho^2}{1-\rho} \left(f\ \frac{f}{\alpha} + (1-f)\ \frac{1-f}{1-\alpha}\right) + \rho.$$

But $f \frac{f}{\alpha} + (1-f) \frac{(1-f)}{1-\alpha} \geq 1$. This can be verified by setting the derivative w.r.t. α of the left hand side equal to zero and observing that this expression has its minimum at $f = \alpha$, at which point it equals one. We conclude that with minimizing the number of people in the system as the criterion, it is possible in some instances that splitting the catalog will improve both the subject and the author-title catalogs as compared to the dictionary catalog, a situation not possible using the other criteria. We remark that this cannot happen if $\alpha = f$, since this reduces the equations to the waiting time forms; thus the possibility of improving both subject and author-title performance depends upon one of the components being underused.

VI. Book Catalog

In this section we consider the implications of our model for the construction of book catalogs. Whereas in a card catalog the manager has little freedom in determining the number of items to be represented in a single drawer, book catalogs offer more freedom by allowing as many pages to be bound in a volume as is desired, and this freedom can be used to control congestion. The question to be considered then, is not whether or not to split the catalog, but rather, into how many volumes to divide the catalog to limit congestion to a desired level.

We now let λ_d be the overall rate of arrivals wishing to use the catalog; of this, $\alpha\lambda_d$ will be for subject use, and $(1-\alpha)\lambda_d$ for author-title use. If we maintain a dictionary catalog in m volumes, the arrival rate per volume will be $\frac{\lambda_d}{m}$; if divided into m_s subject volumes and m_a author-title volumes, the rates will be $\frac{\alpha\lambda_d}{m_s}$ and $\frac{(1-\alpha)\lambda_d}{m_a}$ respectively. It is now possible to use our equations to make assertions for each criterion of congestion.

a) Blocking

Suppose it is desired to keep the probability of blocking below p. Then we must demand $\frac{\lambda t}{m} \leq p$ or $m \geq \frac{\lambda t}{p}$. Thus:

1) $$m_s = \frac{\alpha\lambda_d t_s}{p} ,$$

2) $$m_a = \frac{(1-\alpha)\lambda_d t_a}{p} , \text{ and}$$

3) $$m_d = \frac{\lambda_d[\alpha t_s + (1-\alpha)t_a]}{p} .$$

We note that $m_d = m_a + m_s$, so there is no advantage in keeping the book catalog split or merged if it is desired to keep the total number of volumes to a minimum.

It might be of interest to estimate the relative number of pages in the subject and title-author volumes. If the total number of pages to be bound is N, then fN will be in the subject part and (1-f)N in the author-title part. This yields

$$\frac{(1-f)\ Np}{(1-\alpha)\lambda_d t_a} \quad \text{and} \quad \frac{f\ N\ p}{\alpha\ \lambda\ t_s} ,$$

for the number of pages in the author-title and subject volumes respectively. The ratio of the number of pages in a subject volume, p_s, to that of an author-title volume, p_a, is thus $\frac{p_s}{p_a} = \frac{f}{1-f}\ \frac{1-\alpha}{\alpha}\ \frac{t_a}{t_s}$. We see that the relative thickness of a subject volume is inversely proportional to the average time required for its use. Although the exact number of pages in a volume depends upon the maximum acceptable probability of blocking, p, the ratio of pages does not; if a subject volume has 1/2 the number of pages as does

an author-title volume for one value of p, it will have this property for all values.

b) Waiting Time

We here demand that

$$W \leq \frac{\frac{\lambda v}{m}}{2[1-\frac{\lambda}{m}t]} = \frac{\lambda v}{2[m-\lambda t]} ,$$

where W is the maximum acceptable waiting time. Solving for m and using $v = 2\kappa t^2$, we find

$$m \geq \frac{\lambda v}{2W} + \lambda t = \lambda t(1 + \frac{t\kappa}{W}).$$

For the particular case being considered this becomes

$$m_s \geq \alpha\lambda_d t_s(1 + \frac{t_s\kappa_s}{W}) ,$$

$$m_a \geq (1-\alpha)\lambda_d t_s(1 + \frac{t_a\kappa_a}{W}) , \text{ and}$$

$$m_d \geq \lambda_d[\frac{\alpha\kappa_s t_s^2}{W} + \frac{(1-\alpha)\kappa_a t_a^2}{W}] + \lambda_d[\alpha t_s + (1-\alpha)t_a].$$

We see once again that $m_d = m_a + m_s$. The relative thickness is now given by

$$\frac{p_s}{p_a} = \frac{f}{1-f} \frac{1-\alpha}{\alpha} \frac{\kappa_a t_a^2}{\kappa_s t_s^2} \frac{[1 + \frac{W}{\kappa_a t_a}]}{[1 + \frac{W}{\kappa_s t_s}]}$$

We now have the number of pages in the subject volume essentially decreasing as the square of the time for a use. The κ's have the same effect as increasing the time required for a use.

c) Number in the System

It is now desired to keep the expected number of people demanding a volume below n. The criterion on the number of volumes becomes

$$n \lesssim \frac{\frac{\lambda^2}{m^2} v}{2(1 - \frac{\lambda}{m} t)} + \frac{\lambda}{m} t = \frac{\lambda^2 \kappa t^2}{m^2 - \lambda m t} + \frac{\lambda}{m} t .$$

This produces the quadratic inequality:

$$m^2 - \lambda t(1 + \frac{1}{n})m + \frac{\lambda^2 t^2}{n}(1-\kappa) \leq 0$$

We solve the equation, getting:

$$m = \frac{\lambda t}{2} \left[(1 + \frac{1}{n}) \pm \sqrt{(1 + \frac{1}{n})^2 + \frac{4}{n}(\kappa - 1)} \right]$$

If we expand the solution to first order in $\kappa-1$, we get:

$$m = \frac{\lambda t}{2} \left[(1 + \frac{1}{n}) \pm (1 + \frac{1}{n} + 2(\kappa-1)\frac{n}{(n+1)^2}) \right].$$

Thus either $m \leq \lambda t(1-\kappa) \frac{n}{(n+1)^2}$, or $m \geq \lambda t(1 + \frac{1}{n} + (\kappa-1) \frac{n}{(n+1)^2})$.

The first solution has to be rejected and arises from the formal possibility of making n small by making it negative. But $1 - \frac{\lambda}{m} t > 0$, so the solution is not valid. We note that for $\kappa_a \neq \kappa_s$, it is no longer true that $m_s + m_a = m_d$. The effect of large κ is to increase the number of books required, small κ reducing it. We will, however, anticipate the possibility that $\kappa \approx 1$, and thus write the equations as

$$m_s = \alpha\lambda t_s(1 + \frac{1}{n}) \ ,$$

$$m_a = (1 - \alpha)\lambda t_a(1 + \frac{1}{n}) \ , \text{ and}$$

$$m_d = m_s + m_a .$$

The thickness ratio is now $\frac{p_s}{p_a} = \frac{f}{1-f} \frac{(1 - \alpha)}{\alpha} \frac{t_a}{t_s}$, agreeing with case a).

All the equations express m as the product of λt and a "form factor." λt is the number of arrivals expected in the time required for a single use. Thus, if m were set equal to λt, we would have one volume for each person arriving during a single use. The form factor modifies this number to fit it to the desired behavior -- it expresses the number of volumes required for each person arriving during a use.

We finally note that for a fixed m, $n = \frac{W}{t} = \frac{p}{1-p}$, which allows a comparison, for $\kappa = 1$, of the values of the various criteria.

VII. Conclusion

The problem of congestion at a catalog is much more complex than may have been recognized. We presented three criteria of congestion and found that it is possible that each lead to a different conclusion. To decide which to use may pose some difficulty. Regarding the book catalog, the manager may make estimates for volumes required on the basis of all three criteria and choose the maximum of these estimates, since this will automatically satisfy all conditions.

The question of splitting the card catalog as expressed in its full generality is more difficult to answer. It is likely however, that only one form of congestion is

creating any difficulty and this may be the only one that need be considered; it would be possible, however, to estimate the other forms of congestion before changing the form of the catalog to make certain that the new catalog not introduce problems previously absent. In a functioning library, ρ_a and ρ_s will be quite small and thus blocking and waiting times will not pose a serious problem. However, it may be felt that too many people are in the catalog area. In this case one could examine the number of people to be expected at each type of catalog. Since ρ is small, it is easier to use an approximation to the full equation. Neglecting terms of order of ρ^3 we have

$$n_s = \kappa_s \rho_s^2 + \rho_s,$$

$$n_a = \kappa_a \rho_a^2 + \rho_a,$$

$$n_d = \kappa_s f \frac{f}{\alpha} \rho_s^2 + (1-f)\kappa_a \frac{(1-f)}{(1-\alpha)} \rho_a^2 + f\rho_s + (1-f)\rho_a \text{ , and}$$

$$n_{sp} = \kappa_s f \rho_s^2 + (1-f)\kappa_a \rho_a^2 + f\rho_s + (1-f)\rho_a .$$

With this approximation, the dictionary catalog will be preferred when:

$$(f-\alpha)\left[\kappa_a \frac{(1-f)}{(1-\alpha)} \rho_a^2 - \kappa_s \frac{f}{\alpha} \rho_s^2\right] > 0.$$

This condition is very similar to that obtained when we studied the effect of blocking, and if

$$\kappa_a \frac{(1-f)}{(1-\alpha)} \approx \kappa_s \frac{f}{\alpha} ,$$

the results will agree. (Indeed, if $\kappa_a = \kappa_s$ and $f = \alpha$, conclusions based on all three criteria will agree.) It will be seen, however, that the absolute magnitude of the difference is not likely to be significant.

Consider values that the parameters might take in a heavily used public library. Let $\kappa_s = \kappa_a = 1$, $f = 1/2$, $\alpha = 3/4$, $\rho_s = .2$ and $\rho_a = .01$. In this case, the number of people using each 1000 drawers would be 240 for a subject catalog and 10 for an author-title catalog, or an average of 125 for the split catalog. Should the cards be combined, we would expect 120 people at each 1000 drawers of the dictionary catalog, the difference reflecting a reduction in the number of frustrated users. It is seen that the reduction in the average number of people using the catalog is slight, though the congestion is considerably smaller than that at the isolated subject catalog. In the unlikely event that the ρ's are much larger, the differences would be more dramatic.

We finally note that though we based our paper on the possible splitting of a catalog along subject and author-title lines, our analysis would apply to other forms of splitting as well. Such would be a consideration, for example, should it be contemplated to separate cards for older books from those for more current titles.

Appendix I

$$W_s = \frac{\frac{1}{\lambda_s}\kappa_s \rho_s^2}{1-\rho_s} \quad \text{is of the form } \frac{a}{b},$$

$$W_a = \frac{\frac{1}{\lambda_a}\kappa_a \rho_a^2}{1-\rho_a} \quad \text{is of the form } \frac{c}{d}, \text{ while}$$

$$W_d = \frac{f\frac{1}{\lambda_s}\kappa_s \rho_s^2 + (1-f)\frac{1}{\lambda_a}\kappa_a \rho_a^2}{f(1-\rho_s) + (1-f)(1-\rho_a)} \quad \text{has the form}$$

$$\frac{fa + (1-f)c}{fb + (1-f)d}.$$

Suppose removing author-title cards from a dictionary catalog improves congestion. Then, since the subject catalog remains, we have

$$\frac{a}{b} < \frac{fa + (1-f)c}{fb + (1-f)d}.$$

All terms are greater than zero, so we can multiply both sides by $\frac{b}{a}$ and maintain the sense of the inequality:

$$1 < \frac{f + (1-f)\frac{c}{a}}{f + (1-f)\frac{d}{b}}, \quad \text{or}$$

$$f + (1-f)\frac{d}{b} < f + (1-f)\frac{c}{a}, \text{ which implies}$$

$\frac{d}{b} < \frac{c}{a}$ or $\frac{a}{c} < \frac{b}{d}$. We can now reverse these steps, getting in turn:

$$(1-f) + f\frac{a}{c} < (1-f) + f\frac{b}{d} \quad ,$$

$$1 > \frac{(1-f) + f\frac{a}{c}}{(1-f) + f\frac{b}{d}} \quad \text{, and finally,}$$

$$\frac{c}{d} > \frac{(1-f)c + fa}{(1-f)d + fb} \quad .$$

This implies that if removing the author-title cards from the dictionary catalog improves congestion at the remaining subject catalog, then congestion must be increased at the author-title catalog.

Appendix II

$$W_{sp} < W_d \text{ if } \frac{\frac{\alpha}{\lambda_s}\kappa_s \rho^2 + \frac{1-\alpha}{\lambda_a}\kappa_a \rho^2}{1-\rho} =$$

$$\frac{\rho^2}{(1-\rho)\ \lambda_s\lambda_a}\ [\kappa_a\lambda_s + \alpha(\kappa_s\lambda_a - \kappa_a\lambda_s)]$$

is less than
$$\frac{\frac{f}{\lambda_s}\kappa_s \rho^2 + \frac{1-f}{\lambda_a}\kappa_a \rho^2}{1-\rho} =$$

$$\frac{\rho^2}{(1-\rho)\ \lambda_s\lambda_a}\ [\kappa_a\lambda_s + f(\kappa_s\lambda_a - \kappa_a\lambda_s)] ,$$

or
$$\alpha[\kappa_s\lambda_a - \kappa_a\lambda_s] < f[\kappa_s\lambda_a - \kappa_a\lambda_s] ,$$

or
$$(f-\alpha)\ [\frac{\lambda_a}{\kappa_a} - \frac{\lambda_s}{\kappa_s}] > 0.$$

We note that if $\kappa_a = \kappa_s$, this will always be true, since $\alpha < f$ implies $\lambda_s < \lambda_a$ and $\alpha > f$ implies $\lambda_s > \lambda_a$. If $\kappa_a \neq \kappa_s$, this no longer need be the case; a small κ_s may, for example, effectively increase λ_s so that the inequality may not hold. Since we do expect $\kappa_a \approx \kappa_s$, we hypothesize that for $\rho_s = \rho_a$, a split catalog will always be preferred, unless $\alpha = f$, in which case both are equally acceptable.

Appendix III

If $\alpha = f$, we also have $\lambda_s = \lambda_a$. Then $W_d < W_{sp}$ if

$$\frac{f\ \kappa_s\ \rho_s^2 + (1-f)\ \kappa_a\ \rho_a^2}{f(1-\rho_s) + (1-f)(1-\rho_a)} < \frac{f\ \kappa_s\ \rho_s^2}{1-\rho_s} + \frac{(1-f)\ \kappa_a\ \rho_a^2}{1-\rho_a} .$$

The R.H.S. $= \dfrac{f\ \kappa_s\ \rho_s^2}{1-\rho_s} + \dfrac{(1-f)\ \kappa_a\ \rho_a^2}{1-\rho_s - \Delta} =$

$$\frac{f\ \kappa_s\ \rho_s^2}{1-\rho_s} + \frac{(1-f)\ \kappa_a\ \rho_a^2}{1-\rho_s} \left[1 + \frac{\Delta}{1-\rho_s - \Delta} \right] ,$$

where we define $\Delta = \rho_a - \rho_s$, and use the algebraic identity,

$$\frac{1}{x-y} = \frac{1}{x} \left[1 + \frac{y}{x-y}\right] .$$

Similarly the L.H.S. $= \dfrac{f\ \kappa_s\ \rho_s^2 + (1-f)\ \kappa_a\ \rho_a^2}{1-\rho_s - (1-f)\ \Delta} =$

$$\frac{f\ \kappa_s\ \rho_s^2 + (1-f)\ \kappa_a\ \rho_a^2}{1-\rho_s} \left[1 + \frac{(1-f)\ \Delta}{1-\rho_s - (1-f)\ \Delta} \right] .$$

We use this result to assert $W_d < W_{sp}$ if

$$\frac{(1-f)\ \Delta}{1-\rho_s} \cdot \frac{\kappa_s\ f\ \rho_s^2 + (1-f)\ \kappa_a\ \rho_a^2}{1 - (\rho_s + (1-f)\ \Delta)} < \frac{\Delta\ (1-f)}{1-\rho_s} \quad \frac{\rho_a^2}{1-(\rho_s + \Delta)} ,$$

or 1)

$$\frac{\kappa_s \, f \, \rho_s^2 + (1-f) \, \kappa_a \, \rho_a^2}{1 - \rho_a + f \, \Delta} \, \Delta \; < \; \frac{\kappa_a \, \rho_a^2}{1-\rho_a} \, \Delta \; .$$

We now assume $\kappa_a \rho_a^2 > \kappa_s \rho_s^2$ if $\rho_a > \rho_s$ and v.v. If so, note:

if $\Delta > 0$, $\frac{1}{1-\rho_a + f \, \Delta} < \frac{1}{1-\rho_a}$ and $f \, \kappa_s \, \rho_s^2 + (1-f) \, \kappa_a \, \rho_a^2 < \kappa_a \, \rho_a^2$.

So, multiplying both sides of the inequalities by each other and then both sides by Δ, we obtain 1).

If $\Delta < 0$, the inequalities are reversed, but multiplying these together and then multiplying by $\Delta < 0$ again yields 1). Thus here we always obtain $W_d < W_{sp}$.

Footnotes

1) In some situations other comparisons may prove more fruitful, e.g. the amount of congestion at a dictionary catalog as compared to the congestion at that component of the split catalog suffering the most from congestion. The comparison chosen here should then be viewed only as illustratively the technique used, and not as prescriptive.

2) Further discussion of this problem in the general context of operations research in libraries may be found in (6).

References

(1) James Wilson McGregor. "In Defense of the Dictionary Catalog," Library Resources and Technical Services, vol. 15, no. 1 (Winter, 1971), p. 28.

(2) Fred Heinritz. "Does Dividing the Catalog Relieve Congestion?" Library Resources and Technical Services, vol. 8, no. 3 (September, 1964), p. 310.

(3) Seymour Lubetsky. "Crisis in the Catalog." Catalogers and Classifiers' Yearbook, vol. 9 (1939), p. 48-54.

(4) Thomas L. Saaty. Elements of Queuing Theory, New York: McGraw-Hill, 1961.

(5) Philip M. Morse. Library Effectiveness: A Systems Approach, M.I.T. Press, 1968.

(6) Bookstein, Abraham. (To be published in Library Quarterly, January, 1971.)